Public Disclosure Authorized

POVERTY & EQUITY NOTES

ISSUE 45 | APRIL 19, 2022

PAVING THE WAY TO EQUAL ACCESS: GEOSPATIAL ANALYTICS CAN ADDRESS GEOGRAPHIC DISPARITIES IN SERVICE ACCESSIBILITY IN PAKISTAN

Robert Banick, Manish Basnet, Lander SMM Bosch, and Moritz Meyer[1]

Public Disclosure Authorized

22.8 million Pakistani children aged 5 to 16 do not attend school,[2] the second-largest out-of-school population of any country. Ensuring access to high-quality education and raising school enrollment therefore is the first pillar of Pakistan's World Bank Country Partnership Framework (CPF).[3] This is particularly important for girls, 55% of the out-of-school population.[2] Aside from teacher shortages, distance to school is the largest barrier to accessing education in rural areas, with 29% of parents stating schools are too far away.[4] As a consequence, non-attendance is 9% higher for boys in rural districts (28%) compared to urban areas (19%).[5] For girls, this contrast is even starker. They are twice as likely to not attend school in rural areas compared to their urban peers, with an out-of-school rate of 43% versus 21%. This gap widens for higher education levels. Long commutes are also time- and resource-intensive, and pose safety and security concerns, especially for girls.[6]

While addressing these well-known disparities is crucial, they are rarely measured at the administrative level where service provision happens, such as tehsils (ADM3) in Pakistan. The Transport and Poverty & Equity GPs therefore partnered to apply a methodology to calculate spatial disparities in education accessibility for small administrative units. This allows us to identify priority areas for school provision and commuting roads for upgrading in Khyber Pakhtunkhwa (KP) Province.

1 Authors listed in alphabetical order, roles on last page.

2 NEMIS-AEPAM (2018). Pakistan Education Statistics 2016–17. National Education Management Information System Academy of Educational Planning and Management, Ministry of Federal Education and Professional Training, Government of Pakistan. AEPAM Publication No. 281.

3 World Bank Group (2021). World Bank Group Consultations in Pakistan. Last accessed: January 28th, 2022.

4 Pakistan Bureau of Statistics (2016). Pakistan Social and Living Standards Measurement Survey (PSLM) 2014-15 Provincial/District. Last accessed: March 3rd, 2022.

5 Pakistan Bureau of Statistics (2021). Pakistan Social and Living Standards Measurement Survey (PSLM) 2019-20 Provincial/District. Last accessed: January 28th, 2022.

6 Human Rights Watch (2018). Barriers to Girls' Education in Pakistan. Last accessed: January 31st, 2022.

WORLD BANK GROUP
Poverty & Equity

This note outlines the methodology, which can be upscaled, replicated and applied to other geographic contexts and a wide diversity of services, with the help of the publicly available toolkit stemming from this project. Current access conditions can be quantified, and improvements in accessibility through the rehabilitation or construction of specific roads simulated. Data-driven investment decisions and policy can then be optimized regarding the location and mix of new services or transportation upgrades.

INTRODUCTION

Detailed spatial knowledge of service accessibility disparities is crucial to design well-targeted policies and projects to address them, both for the World Bank Group and its clients. However, data aiming to showcase spatial disparities in service access tend to be derived from surveys focusing on large geographic areas or case studies with a limited geographic scope. This obfuscates insights into areas and communities that are disadvantaged at finer scales. Where this granularity is sought, approximative modelling methods have traditionally been applied, such as counting populations within buffers of fixed, Euclidian distances around services or roads. While physical accessibility is in first instance determined by the location of services and the roads connecting them with communities, many geographic determinants complicate simple linear distance-based models. *As-the-crow-flies* methods vastly underestimate travel distances along road networks, and by not accounting for the terrain and road conditions, they are unable to accurately predict travel times.[7] The lack of differentiation in modes of transport further reduces their predictive power. Walking along an icy mountain trail will be slower than driving along a paved highway on a plain, with significant repercussions on the access of communities to services and opportunities. Approximative techniques that do not move beyond distance will therefore only deliver approximative results. They are unable to accurately quantify and visualise spatial disparities in service provision, particularly where terrain or poor infrastructure complicate travel.[8] Accessibility models must account for terrain gradients, altitude, seasonality, built characteristics, vegetation, and transport modes.

Remoteness and limited access to services and opportunities is systematically linked to worse development outcomes through adverse human capital

[7] Shahid, R. *et al.* (2009) Comparison of distance measures in spatial analytical modeling for health service planning. BMC Health Services Research 9(200),

[8] Banick, R. & Kawasoe, Y. (2019) Measuring Inequality of Access: Modeling Physical Remoteness in Nepal. Policy Research Working Paper; No. 8966. World Bank, Washington, DC.

outcomes and fewer opportunities to generate income.[9] Data on inequalities in socioeconomic and human development, presented at aggregated levels from surveys, also lack the necessary granularity to inform policy and project design. Hence, vulnerable groups risk being left behind when policies are made, suffering poorer educational, health and socioeconomic outcomes.

Confronted with strong disparities in children's access to education observed across KP Province but lacking the granular data to identify the most deprived communities and areas, the **Pakistan Poverty & Equity team refined and applied a method to measure and visualise accessibility disparities with high spatial resolution.** This work is done in collaboration with the Transport team, seeking to inform data-driven investment decisions around the location of new schools and transportation upgrades in KP Province, especially to benefit girls' education. It builds on and refines previous accessibility work performed in Afghanistan,[10] Nepal,[11] Sri Lanka,[12] Bhutan, and Bangladesh.[13]

The resulting freely and openly available methodological toolkit[14] answers to the need for accurate calculations of current access conditions, and the projection of potential improvements in accessibility for specific roads and small areas. This method is replicable in other geographic contexts for a wide variety of services.

DATA & METHODOLOGY

The multidimensionality of accessibility inherently requires a broad set of data inputs. Models are only as good as the data they are built on. The overview of required data inputs for accessibility analyses is listed in *Table 1*, substantiated by the data used by the team working on accessibility to services in KP Province. High-quality, country- or province-specific datasets are recommended for use where available. Significant time should be allocated to sourcing and validating reliable spatial datasets from government partners and development peers. Where data gaps exist, public data, such as OpenStreetMap (OSM) data, can be used in addition

[9] UN DESA (2020) World social report 2020 - Inequality in a rapidly changing world. UN publication, ISBN 978-92-1-130392-6.

[10] World Bank (2020) How To Get There? In Afghanistan, It's As Simple As Putting 'There' On The Map. Last accessed: February 3rd, 2022.

[11] Heyns, A.; Banick, R. & Regmi, S. (2021) Roads Development Optimization for All-Season Service Accessibility Improvement in Rural Nepal Using a Novel Cost-Time Model and Evolutionary Algorithm. Policy Research Working Paper; No. 9526. World Bank, Washington, DC.

[12] World Bank Data Development Hub (2021) Model and Results For Optimizing Tourism Investments: Identifying Chokepoints and Underperforming Minor Roads In Sri Lanka. Last accessed: February 11th, 2022.

[13] World Bank (2022) Cox's Bazar - Inclusive Growth Diagnostic. World Bank, Washington, DC.

[14] Toolkit available at GitHub. Last accessed: February 11th, 2022.

to manual data collection. Geocoded road, settlement, school location (or other points of interest) and population data need to be complemented with raster-based terrain and environmental data, and estimations of travel speeds. To obtain policy and project-relevant information, the results can be computed and visualized for small administrative units, requiring official and up-to-date boundary delineations.

Table 1. Data inputs for accessibility modelling

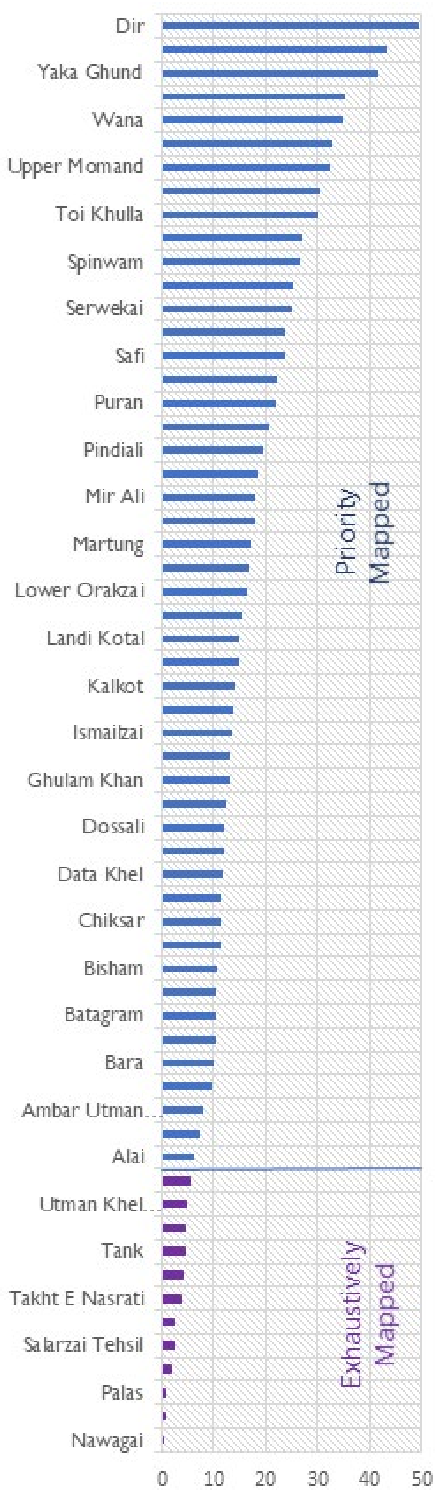
Theme	Dataset required	Example for KP Province, Pakistan
Roads	Vector-based road network and attributes (condition, surface, type, etc.)	OSM, Pakistan Communication and Works Department, manual verification
Terrain	Elevation & Slope, high-resolution (30m) raster	NASA Shuttle Radar Topology Mission 2000 (released in 2015)
Terrain	Landcover, high-resolution (30m) raster	ICIMOD 2010; FAO 2021 Land Cover Classification
Environment	Mean elevation data (see above), season lengths	NASA SRTM 2000, World Weather Online
Speed	Seasonal and road-type walking/vehicle speeds	Irmischer & Clarke 2018; WB Poverty & Transport GP
Services	Geocoded school locations	MSI Pakistan/USAID KP school data
Population	Population model, high-resolution (30m) raster	Facebook/CIESIN High Resolution Satellite Layer 2015 (good at identifying uninhabited areas) or WorldPop 2020 (good at population distribution in inhabited areas)
Settlement	Settlement model, high-resolution (10m) raster	DLR World Settlement Footprint, 2015
Administrative	Administrative boundary shapefiles Admin 0-3	UN-OCHA 2018
Welfare	Development outcome indicators at small administrative levels	Government of Pakistan, Pakistan Social And Living Standards Measurement Survey

Road data form the backbone of sound accessibility models. Their completeness and quality are critical to model accuracy. Unfortunately, in South Asia neither governmental nor open data roads datasets tend to be fully complete, particularly in rural areas. While governmental roads databases have the advantage of being official and are frequently linked to useful maintenance data, these datasets might not be accessible to the public or collaborating peers. They might also be incomplete or outdated, particularly if different agencies manage separate parts of the road network. In KP Province, for instance, smaller rural roads were entirely absent from the administrative data. OSM offers a freely accessible open-source alternative which tends to provide better coverage than Government sources in South Asia, especially for smaller, rural roads. However, the coverage extent and data quality vary unpredictably across space, and relevant road attributes like the surface material are not always consistently recorded. OSM data also tend to be less readily accepted by Government counterparts.

Crucially, OSM data can be quickly improved using freely available satellite imagery and editing software maintained by OSM's active volunteer community. In Pakistan, the analytical team, together with consultants and volunteers, mapped missing roads, and reclassified those who were assigned the wrong category in OSM (motorway, trunk, primary, secondary, tertiary, service, residential road, or unpaved rural roads) using freely available satellite imagery. Not only does this deliver the best available road networks for accessibility modelling, the OSM data created are also freely available to others working on the area of study. The cost per square kilometre for upgrading OSM data depends on the quality of satellite imagery and road density. Manually digitizing each road is, however, time- and resource-intensive for large areas, especially if densely populated. To reduce costs, we trialled a prioritization method that allows to identify populated areas not adequately covered by the OSM road network. This is done by filtering World Settlement Footprint rasters to remove minor hamlets and then identifying settled areas more than 500 meters from an existing road in OSM. These major unconnected populated areas were then manually integrated into the existing OSM road network by identifying and digitizing the main connecting road.

Improving OSM substantially improves the accuracy of accessibility modelling, and a prioritized approach can yield these improvements at a fraction of the cost and time of an exhaustive approach. Mapping missing roads was found to change aggregate accessibility figures per *tehsil* up to 50%, resulting in meaningful shifts in the overall rankings of *tehsils* (*Figure 1*). This proves crucial in delivering accurate outcomes, as three out of the top-ten least accessible *tehsils* requiring road and service investments would not have been identified without these improvements. Accessibility improvements from the priority mapping were found to be greatest in rural areas, where OSM data tended to be less complete. The prioritization method proved equally valuable in flat, hilly, and mountainous *tehsils*. Priority-only mapping was found to cut costs fourfold and reduce mapping time by half compared to exhaustively mapping each missing road in the OSM network for KP Province, Pakistan. While minor further changes in case of exhaustive mapping cannot be ruled out, the additional mapping time and cost would return significantly decreased marginal benefits. It is important to note that improving OSM data through reclassification of road categories may also entail negative changes in accessibility figures. This happens when roads classified as well-maintained or allowing for fast travel speeds in OSM are of poorer quality or hamper high speeds in reality. Correcting these errors then allocates lower speeds to these roads, effectively reducing accessibility in the model to accurately reflect the situation on the ground.

Figure 1: Improvements (%) in accessibility to services through the additional OSM mapping exercise for KP *Tehsils*, broken down by priority and full-scale mapping



Having improved OSM data, we combined these with governmental data to get the best possible integrated dataset. Both datasets were aligned to a simplified data model representing roads in terms of modelled travel speeds. Four road classes and surface types and three approximate levels of road curvature and condition were created. This involves simplifying any number of input data attributes to those of our model. For instance, asphalt, bitumen, and concrete are all represented as "paved" surfaces. Where surfaces or conditions were missing in a dataset, we imputed them based on the class of a road. A matrix of these factors determines the base speed for each road segment and the degree to which anticipated seasonal effects (ice, mud, etc.) further reduce speeds for seasonal models. The combined terrain, environment, and road data components represent a certain 'friction', slowing down or speeding up travel.

Integrating all these characteristics in high-resolution, 30-metre resolution cells, a granular 'friction surface' can be created. The friction surface represents the time it takes to cross each cell. Underlying this raster-based surface is, first, the average on-network travel speeds over transportation links. These are calculated based on the road class, surface material, estimated condition, and curvature in the data model described above. Baseline speeds are calculated using Government road standards validated by World Bank Transport Specialists. Next, off-network walking speeds are computed based on the slope of underlying terrain and landcover (*Figure 2a*)[15]. These speeds are modified by environmental conditions, including seasonal speed reductions due to mud, ice or snow and slower walking speeds at high altitudes where lower oxygen levels reduce aerobic capacity. Water bodies are modelled to block passage, except where bridges are present. For each cell of the friction surface, we estimate how long it takes to cross that cell using the fastest available means of transport – walking off-road, or by vehicle on-road, with the appropriate speed (*Figure 2b*).[15] Linking the exact location of services and settlements to their related cells on the friction surface, travel times between settlements and points of interest can then be calculated (*Figure 2c*).[15]

Friction maps, which differ by season (dry, monsoon and winter), are then overlayed with population models. Population models show the distribution of human population at the same spatial resolution as the friction cells.[16] Multiplying travel times with the population living in each cell yields the 'person-hours' per grid cell (*Figure 2d*).[15] This measure can yield the population-weighted accessibility to a service per administrative unit by aggregating the accessibility of the population living in cells that make up the relevant spatial units. Highly detailed gaps in service provision can then be spatially identified – in effect, the only limit on aggregation is the availability of spatial unit boundaries. Linking this picture to geographic disparities in development outcomes, the compounding effects of deprivation caused by low access to services can be brought to light. Pinpointing neatly delineated areas and communities that suffer from unequal access to services builds the case for the targeted provision of services and/or the upgrading or construction of roads and transport infrastructure to allow levelling up across the country.

Beyond identifying spatial disparities, the method enables simulations of the relative improvements in service accessibility due to hypothetical road upgrades (*Figure 2e*).[15] An upgraded or new road will increase travel speeds in the cells of the friction surface it crosses, lowering the travel time between homes and services. Subtracting the real-life 'before' and simulated 'after' access figures returns the time-wise access improvements for each investment in upgrading or extending the road network, indicating where this is most beneficial for the population.

Given their technical complexity, computational intensity, and reliance on context-specific parameters, raster models are relatively time-consuming to create and can be difficult to communicate. Consequently, they are most effective and beneficial where terrain is a major factor, off-road travel is common, and/or transport network data are likely incomplete. To facilitate their development, the **open-source Python toolkit for raster-based accessibility models and investment statistics** developed for the KP project is now public[14]. This code complements the GOSTNets toolkit[17] for vector analysis developed by the World Bank's Geospatial Operations

[15] Banick, R.; Heyns, A.M. & Regmi, S. (2021) Evaluation of rural roads construction alternatives according to seasonal service accessibility improvement using a novel multi-modal cost-time model: A study in Nepal's remote and mountainous Karnali province. Journal of Transport Geography 93(103057).

[16] Kilic, T. et al. (2016) A first look at Facebook's high-resolution population maps. Last accessed: February 3rd, 2022.

[17] Available at GitHub - GOSTnets. Last accessed: February 23rd, 2022.

Support Team (GOST) and applies tools developed in the GOSTNets Raster[18] at scale within a comprehensive workflow.

Figure 2a. Calculating off-road speeds based on slope and landcover

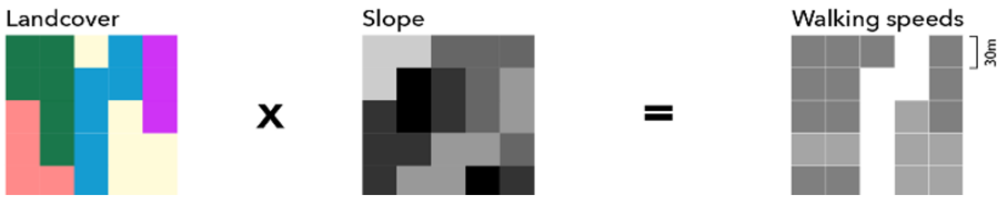


Figure 2b. Calculating the fastest mesh speed from on- and off-road speeds

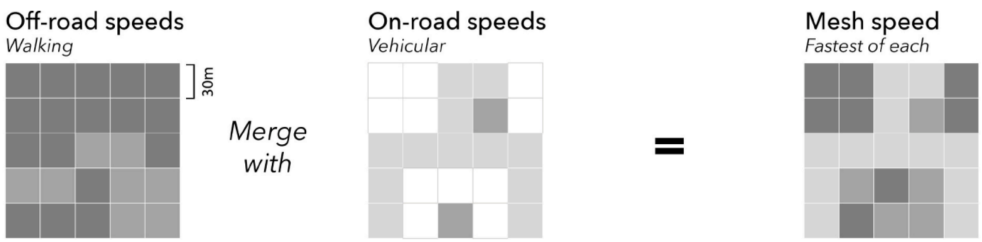


Figure 2c. Calculating access times through friction surfaces

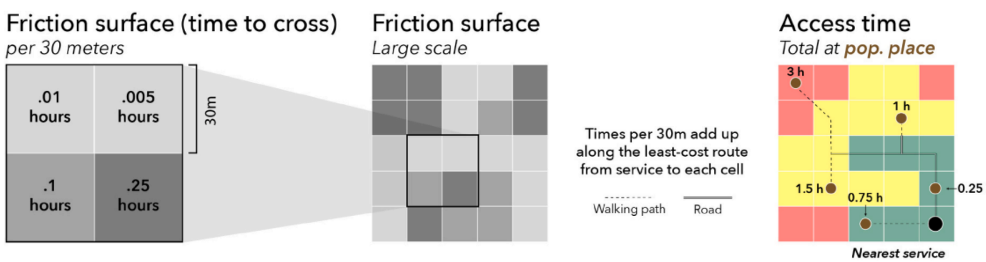
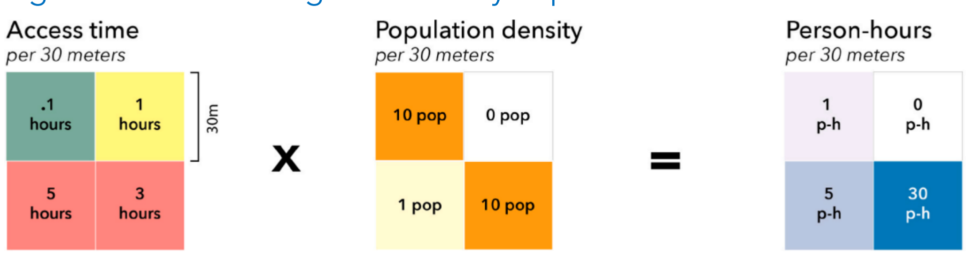


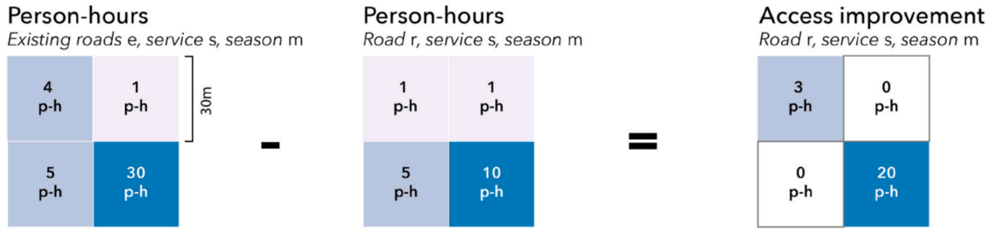
Figure 2d. Calculating accessibility in person-hours



---

[18] Available at GitHub – GOSTNets Raster. Last accessed: February 23rd, 2022.

Figure 2e. Calculating access improvement potential



Source: World Bank, Pakistan Poverty and Equity Team.

Despite the tremendous opportunities these accessibility models entail, challenges and limitations remain. In KP Province, only public service data were available. This implies that the sizeable private education sector could not be included in the analyses. Speeds and road conditions would also benefit from ground truthing for a more accurate assessment of travel time, which was not possible for these models. Looking ahead, the models can also be further developed to focus on the ideal placement of additional services, in addition to the road upgrade work presented in this note, to provide an even more comprehensive picture of pathways to improving service access.

RESULTS

The Pakistan Poverty & Equity and Transport teams set out to identify spatial disparities in public school access to find the largest groups of students suffering from the lowest accessibility to education in KP Province. A province-wide friction surface with a 30-metre resolution was developed, overlaid with a detailed population grid, and person-hour travel times to primary, middle, and high school calculated. These high-resolution raster data were aggregated at the smallest formally recognized administrative level, *tehsils,* classified as Administrative 3-level spatial units. Accessibility values were indexed, normalised, and quintiles calculated. This approach delivers a detailed image of spatial disparities in access to education across KP Province for small administrative units.

***Tehsils* in the mountainous north and rural southwest of the province have the lowest accessibility to all school levels** (*Figure 3a*). Despite these *tehsils* being less populated compared those around the province capital, Peshawar, the population-weighted index of accessibility points to the most pressing need to increase education accessibility in those areas. Children living in the southwest and north of KP Province are thus confronted with structural inequalities in learning opportunities due to limited access to primary, secondary and high schools. As second-order priority, also those *tehsils* along the Afghan border require the attention of policymakers and planners.

In line with the Pakistan CPF, a specific focus should lie on the improvement of girls' access to education. Breaking down analytics for boys and girls, and male and female schools, gendered access disparities were computed. The mountainous north of the province again shows the strongest disparities (*Figure 3b*). Education accessibility in these *tehsils* lies some 20% to 50% lower for girls than for boys. Compared to an equal-access scenario on the 45-degree line, girls consistently have longer travel times to school (*Figure 4*). In areas where boys travel for 20 minutes on average, girls' commuting time is 50% longer – a difference increasing with longer overall commutes and growing by schooling level.

The accessibility analyses thus provide a substantial evidence base for the investment in education in the north of the province, especially benefitting girls in line with the CPF.

Figure 3a. Accessibility quintiles in KP Province

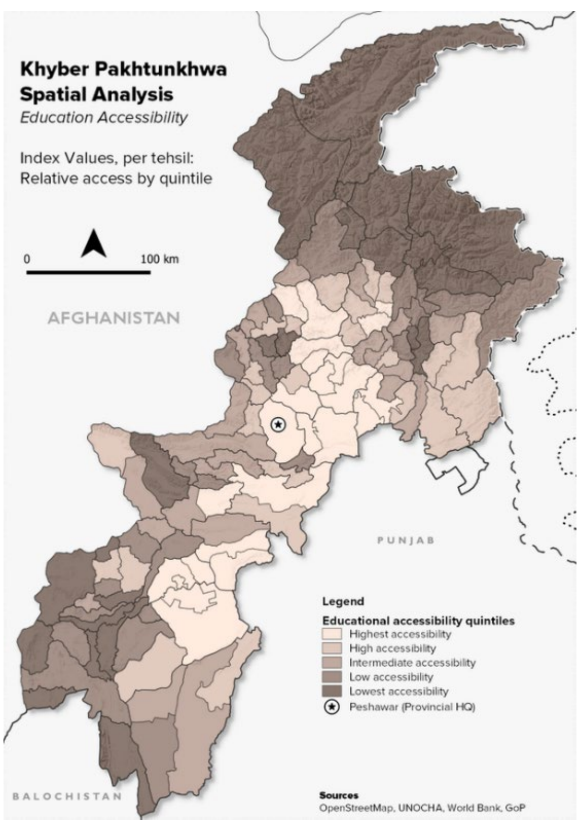
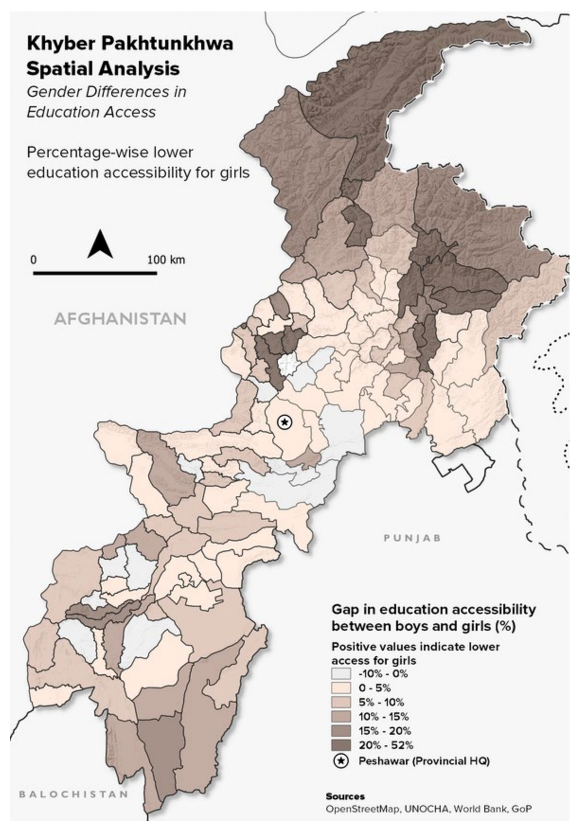
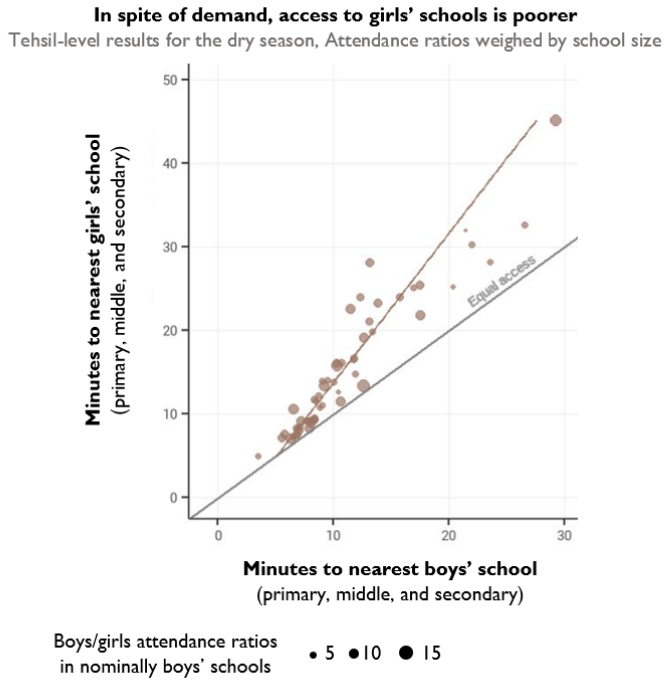


Figure 3b. Gender disparities in access to education in KP Province



Source: World Bank, Pakistan Poverty and Equity Team.

Figure 4. Additional travel time to school for girls in KP Province

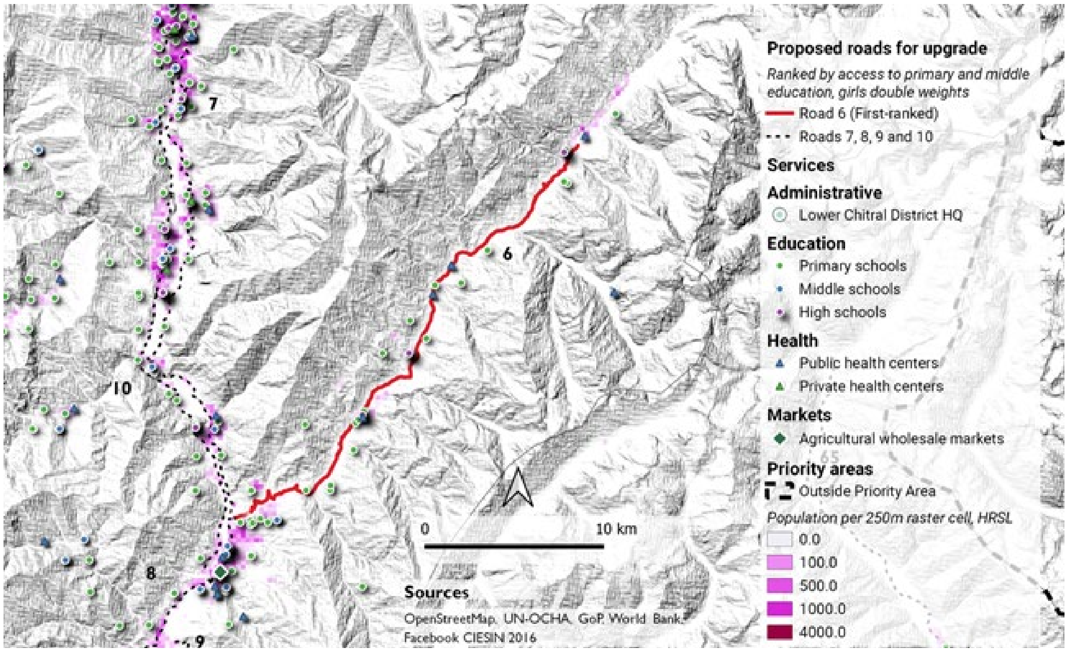


Source: World Bank, Pakistan Poverty and Equity Team.

In addition to identifying small areas where new services can be provided to improve accessibility, access can also be improved through upgrades to the road network. Exploring this option in KP Province, the Provincial Government proposed 79 roads that could be considered for upgrade through the Bank-funded KP Rural Accessibility Project (KPRAP), which is currently being designed. Applying the accessibility methodology to the proposed roads, and comparing current conditions with hypothetical post-investment scenarios, we computed which road investments would offer the highest improvements in accessibility relative to the status quo.

Consistent with the Tehsil-level findings, roads located in the districts with the lowest accessibility and highest gender disparities would provide the greatest improvements in accessibility if upgraded. An example is shown below for 'Road 6', the rural road holding the greatest potential for access improvements, located in Chitral *tehsil* (*Figure 5*). This evidence can be translated into policy and steer investment decisions through KPRAP, directly benefiting the accessibility to education for children, and especially girls, commuting to school in deprived *tehsils*.

Figure 5. 'Road 6', the rural road with highest improvement potential for access to education in KP Province



Source: World Bank, Pakistan Poverty and Equity Team.

POLICY IMPLICATIONS

The World Bank has a long history of innovative work in the field of cost-distance models and measuring access. Unsurprisingly, accessibility analyses have rapidly become a key building block of the Bank's service provision analytics. They provide the essential evidence base informing Government clients, stakeholders, and Bank operations on where gaps in service access are most pressing, and where the potential for improved accessibility is greatest. As such, innovative spatial models of accessibility that go beyond approximate techniques are critical to enable data-driven service and infrastructure investment decisions and to design coherent policies to close spatial disparities in service accessibility. Connecting these spatial insights with the study of welfare disparities also allows deprived communities to level up and reduce inequalities in socioeconomic and human development outcomes.

The methodology developed and applied in KP Province allows to measure accessibility at a highly granular level for remote areas where modelling was previously highly challenging. With limited funds and significant heterogeneity in access and welfare outcomes, disparities should be studied at the finest scale possible. Starting from highly granular rasters, the methodology enables the identification of spatial patterns for the smallest possible administrative units, supporting a focus of funds and human resources on those areas and communities that need it most, accounting for population density.

Accessibility models can assist in closing the service access and usage gap in areas with high service demand where constraints are primarily physical. However, compounding socioeconomic factors limiting access need to be identified and addressed equally effectively. For education, this includes social norms around girls' education[6], the cost of school transportation, teacher shortages, and the quality of education[4]. Tackling physical accessibility disparities is but one piece of the puzzle, and service usage might not increase if remaining critical constraints are not efficiently targeted.

Accessibility analyses like the ones for KP Province are replicable and scalable to other geographic contexts. Accessibility can be determined to any service or opportunity that has spatial characteristics, including healthcare, education, administrative services, and food markets. Moreover, the analyses can be performed anywhere the set of road, population, terrain, and environmental data is available. A highly flexible tool, it has the potential to become a prime resource in identifying and addressing accessibility disparities around the globe, and our freely available Python toolkit can support this work.

About The Authors

Robert Banick worked with the World Bank's Poverty and Equity Global Practice as a geospatial consultant.

Manish Basnet is a consultant working with the World Bank's Poverty and Equity Global Practice on the Bhutan, Nepal, and Pakistan Programs.

Lander SMM Bosch is the Regional Geographer for the South Asia Region in the World Bank's Poverty and Equity Global Practice and a World Bank Young Professional.

Moritz Meyer is Senior Economist in the World Bank's Poverty and Equity Global Practice and Task Team Leader of the Pakistan Poverty and Equity team.

CONNECT WITH POVERTY & EQUITY GLOBAL PRACTICE

www.worldbank.org/poverty @WBG_Poverty

This note series is intended to summarize good practices and key policy findings on Poverty-related topics. The views expressed in the notes are those of the authors and do not necessarily reflect those of the World Bank, its board, or its member countries. Copies of the notes from this series are available on worldbank.org/poverty.